Daniel C. Ferreira

daniel.ferreira.1@gmail.com
github.com/dcferreira | www.dcferreira.com
Based in Vienna, Austria | Born in 1992

Highlights

- → Around 8 years of experience in Machine Learning (industry & academia)
- → Strong mathematical background
- → Curious and fast learner
- → Most interested in NLP and cybersecurity topics

Work Experience

2019–2022 Machine Learning Engineer at Cyan Security

– The position involved a mix of ML engineering, data engineering, MLOps, and research. Topics included website categorization (using multilingual text and images), DNS tunneling detection, and IoT security. Heavy emphasis on developing production-ready ML models, and following MLOps best practices. Using state-of-the-art Machine Learning models (e.g., Transformers).

2016–2019 Researcher in the Communications Networks group, TU Wien, integrated in the Big-DAMA project.

– Research topics mostly related to how to represent network traffic for detecting attacks at the network level (i.e., which features to use). Experimented with both classical features, and feature learning approaches using Deep Learning techniques, and in particular representing traffic in 2-dimensional spaces.

2016 Junior Researcher at Priberam Labs, integrated in the SUMMA project.

– Worked in a massive project for automating media monitoring. In particular, tackled a *named entity recognition* problem, using both classical methods, and Deep Learning approaches for NLP.

Education

2015 MSc in Applied Mathematics at **Instituto Superior Técnico**
Major: Computation
Thesis in cross-lingual classification (grade: 19/20)

– I wanted to classify news documents in German, given an English-only training set. I proposed two novel approaches to find word representations in a bilingual space, one using CCA and another using an original formulation. Details can be found here.

Advisors: André F. T. Martins, Mariana S. C. Almeida, M. Rosário Oliveira

2013 BSc in Applied Mathematics and Computation at **Instituto Superior Técnico**

Highlighted Projects

- → Toxic News: website enabling automatic ranking of online media outlets using machine learning models, based on toxicity/sentiment of the headlines
- → Tweet Fake: generate tweets using LLMs conditioned on the user's style
- → City-GAN: generate images of urban facades, stylized according to the chosen city
- → Traffic Flow Mapper: a prototype to visualize network traffic flows in real time in a 2D plot
- → mdcgenpy: a generator of random clustered data, for evaluation of clustering algorithms
- → Heavily contributed to DeepArchitect: a framework for neural network architecture search
- → multilingual word embeddings: one of the first publicly available multilingual embeddings

Highlighted Publications

2019 Extreme Dimensionality Reduction for Network Attack Visualization with Autoencoders, IJCNN 2019
2019 Towards modular and programmable architecture search, NeurIPS 2019
2016 Jointly Learning to Embed and Predict with Multiple Languages, ACL 2016

Software Experience

Professional proficiency	Python, Linux, Pandas, Transformers, Tensorflow, Keras, scikit-learn, NumPy, Docker, Databricks, Spark, LaTeX, Wireshark
Some professional experience	JavaScript, PyTorch, Scapy, GCP, AWS, Azure, SQL
Hobby experience	MongoDB, R, Go, Rust, C, Photoshop, Inkscape